

[Home](#)

[AI](#)

[Machine Learning](#)

[DBMS](#)

[Java](#)

[Blockchain](#)

[Control System](#)

[Selenium](#)

[HTML](#)

Precision and Recall in Machine Learning

While building any machine learning model, the first thing that comes to our mind is how we can build an accurate & 'good fit' model and what the challenges are that will come during the entire procedure. Precision and Recall are the two most important but confusing concepts in Machine Learning. ***Precision and recall are performance metrics used for pattern recognition and classification in machine learning.*** These concepts are essential to build a perfect machine learning model which gives more precise and accurate results. Some of the models in machine learning require more precision and some model requires more recall. So, it is important to know the balance between Precision and recall or, simply, precision-recall trade-off.



In this article, we will understand Precision and recall, the most confusing but important concepts in machine learning that lots of professionals face during their entire data science & machine learning career. But before starting, first, we need to understand the **confusion matrix** concept. So, let's start with the quick introduction of Confusion Matrix in Machine Learning.

Confusion Matrix in Machine Learning

Confusion Matrix helps us to display the performance of a model or how a model has made its prediction in Machine Learning.

Confusion Matrix helps us to visualize the point where our model gets confused in discriminating two classes. It can be understood well through a 2×2 matrix where the row represents the **actual truth labels**, and the column represents **the predicted labels**.

		Predicted	
		Positive	Negative
Ground-Truth	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

This matrix consists of 4 main elements that show different metrics to count a number of correct and incorrect predictions. Each element has two words either as follows:

- True or False
- Positive or Negative

If the predicted and truth labels match, then the prediction is said to be correct, but when the predicted and truth labels are mismatched, then the prediction is said to be incorrect. Further, positive and negative represents the predicted labels in the matrix.

There are four metrics combinations in the confusion matrix, which are as follows:

- **True Positive:** This combination tells us how many times a model correctly classifies a positive sample as Positive?
- **False Negative:** This combination tells us how many times a model incorrectly classifies a positive sample as Negative?
- **False Positive:** This combination tells us how many times a model incorrectly classifies a negative sample as Positive?
- **True Negative:** This combination tells us how many times a model correctly classifies a negative sample as Negative?

Hence, we can calculate the total of 7 predictions in binary classification problems using a confusion matrix.

Now we can understand the concepts of Precision and Recall.

What is Precision?

Precision is defined as the *ratio of correctly classified positive samples (True Positive) to a total number of classified positive samples* (either correctly or incorrectly).

Precision = True Positive / True Positive + False Positive

Precision = TP / TP + FP

- TP- True Positive
- FP- False Positive
- The precision of a machine learning model will be low when the value of;

TP + FP (denominator) > TP (Numerator)

- The precision of the machine learning model will be high when Value of;

TP (Numerator) > TP + FP (denominator)

Hence, **precision helps us to visualize the reliability of the machine learning model in classifying the model as positive.**

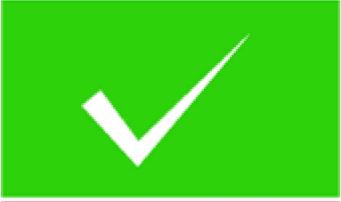

Examples to calculate the Precision in the machine learning model

Below are some examples for calculating Precision in Machine Learning:

Case 1- In the below-mentioned scenario, the model correctly classified two positive samples while incorrectly classified one negative sample as positive. Hence, according to precision formula;

Precision = 0.667

Negative**Positive**

Precision = TP/TP+FP

$$\text{Precision} = 2/2+1 = 2/3 = 0.667$$

Case 2- In this scenario, we have three Positive samples that are correctly classified, and one Negative sample is incorrectly classified.

Put TP =3 and FP =1 in the precision formula, we get;

Precision = 0.75

Negative**Positive**

X	✓
✓	✓
X	✓

Precision = TP/TP+FP

$$\text{Precision} = 3/3+1 = 3/4 = 0.75$$

Case 3- In this scenario, we have three Positive samples that are correctly classified but no Negative sample which is incorrectly classified.

Put TP =3 and FP =0 in precision formula, we get;

Precision = TP/TP+FP

$$\text{Precision} = 3/3+0 = 3/3 = 1$$

Hence, in the last scenario, we have a precision value of 1 or 100% when all positive samples are classified as positive, and there is no any Negative sample that is incorrectly classified.

What is Recall?

The recall is calculated as the ratio between the numbers of Positive samples correctly classified as Positive to the total number of Positive samples. The **recall measures the model's ability to detect positive samples**. The higher the recall, the more positive samples detected.

$$\text{Recall} = \text{True Positive} / \text{True Positive} + \text{False Negative}$$

$$\text{Recall} = \text{TP} / \text{TP} + \text{FN}$$

- TP- True Positive
- FN- False Negative
- Recall of a machine learning model will be low when the value of;
TP+FN (denominator) > TP (Numerator)
- Recall of machine learning model will be high when Value of;
TP (Numerator) > TP+FN (denominator)

Unlike Precision, Recall is independent of the number of negative sample classifications. Further, if the model classifies all positive samples as positive, then Recall will be 1.

Examples to calculate the Recall in the machine learning model

Below are some examples for calculating Recall in machine learning as follows

Example 1- Let's understand the calculation of Recall with four different cases where each case has the same Recall as 0.667 but differs in the classification of negative samples. See how:

Recall = 0.667

Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
X	✓	X	✓	✓	✓	✓	✓
X	✓	✓	✓	✓	✓	✓	✓
X	X	X	X	X	X	✓	X
A		B		C		D	

In this scenario, the classification of the negative sample is different in each case. Case A has two negative samples classified as negative, and case B have two negative samples classified as negative; case c has only one negative sample classified as negative, while case d does not classify any negative sample as negative.

However, recall is independent of how the negative samples are classified in the model; hence, we can neglect negative samples and only calculate all samples that are classified as positive.

Recall=0.667

Negative	Positive	Negative	Positive	Negative	Positive	Negative	Positive
	✓		✓		✓		✓
	✓		✓		✓		✓
	✗		✗		✗		✗
A		B		C		D	

In the above image, we have only two positive samples that are correctly classified as positive while only 1 negative sample that is correctly classified as negative.

Hence, true positivity rate is 2 and while false negativity rate is 1. Then recall will be:

$$\text{Recall} = \text{True Positive} / \text{True Positive} + \text{False Negative}$$

$$\text{Recall} = \text{TP} / \text{TP} + \text{FN}$$

$$= 2 / (2 + 1)$$

$$= 2/3$$

$$= 0.667$$

Note: This means the model has correctly classified only 0.667% of Positive Samples

Example-2

Now, we have another scenario where all positive samples are classified correctly as positive. Hence, the True Positive rate is 3 while the False Negative rate is 0.

Recall=1.0

Negative	Positive
	✓
	✓
	✓

$$\text{Recall} = \text{TP} / \text{TP} + \text{FN} = 3 / (3 + 0) = 3/3 = 1$$

If the recall is 100%, then it tells us the model has detected all positive samples as positive and neglects how all negative samples are classified in the model. However, the model could still have so many samples that are classified as negative but recall just neglect those samples, which results in a high False Positive rate in the model.

Note: This means the model has correctly classified 100% of Positive Samples.

Example-3

In this scenario, the model does not identify any positive sample that is classified as positive. All positive samples are incorrectly classified as Negative. Hence, the true positive rate is 0, and the False Negative rate is 3. Then Recall will be:

$$\text{Recall} = \text{TP} / \text{TP} + \text{FN} = 0 / (0 + 3) = 0/3 = 0$$

This means the model has not correctly classified any Positive Samples.

Difference between Precision and Recall in Machine Learning

Precision	Recall
It helps us to measure the ability to classify positive samples in the model.	It helps us to measure how many positive samples were correctly classified by the ML model.
While calculating the Precision of a model, we should consider both Positive as well as Negative samples that are classified.	While calculating the Recall of a model, we only need all positive samples while all negative samples will be neglected.
When a model classifies most of the positive samples correctly as well as many false-positive samples, then the model is said to be a high recall and low precision model.	When a model classifies a sample as Positive, but it can only classify a few positive samples, then the model is said to be high accuracy, high precision, and low recall model.
The precision of a machine learning model is dependent on both the negative and positive samples.	Recall of a machine learning model is dependent on positive samples and independent of negative samples.
In Precision, we should consider all positive samples that are classified as positive either correctly or incorrectly.	The recall cares about correctly classifying all positive samples. It does not consider if any negative sample is classified as positive.

Why use Precision and Recall in Machine Learning models?

This question is very common among all machine learning engineers and data researchers. The use of Precision and Recall varies according to the type of problem being solved.

- If there is a requirement of classifying all positive as well as Negative samples as Positive, whether they are classified correctly or incorrectly, then use Precision.
- Further, on the other end, if our goal is to detect only all positive samples, then use Recall. Here, we should not care how negative samples are correctly or incorrectly classified the samples.

Conclusion:

In this tutorial, we have discussed various performance metrics such as confusion matrix, Precision, and Recall for binary classification problems of a machine learning model. Also, we have seen various examples to calculate Precision and Recall of a machine learning model and when we should use precision, and when to use Recall.

[← Prev](#)[Next →](#)

 [For Videos Join Our Youtube Channel: Join Now](#)


Feedback

- Send your Feedback to feedback@javatpoint.com

Help Others, Please Share





Learn Latest Tutorials


 [Splunk tutorial](#)
Splunk

 [SPSS tutorial](#)
SPSS

 [Swagger tutorial](#)
Swagger

 [T-SQL tutorial](#)
Transact-SQL

 [Tumblr tutorial](#)
Tumblr

 [React tutorial](#)
ReactJS

 [Regex tutorial](#)
Regex

 [Reinforcement learning tutorial](#)
Reinforcement Learning

 [R Programming tutorial](#)

 [RxJS tutorial](#)
RxJS

 [React Native tutorial](#)

 [Python Design Patterns](#)